UNIVERSITY OF ILLINOIS BULLETIN

Issued Weekly

Vol. XIX July 3, 1922 No. 45

[Entered as second-class matter December 11, 1912, at the post office at Urbana, Illinois, under the Act of August 24, 1912. Acceptance for mailing at the special rate of postage provided for in section 1103, Act of October 3, 1917, authorized July 31, 1918.]

BULLETIN NO. 10

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

RELATION OF SIZE OF CLASS TO SCHOOL EFFICIENCY

by

Bureau of Educational Research

Prepared, in part, from a report by P. R. Stevenson, former Assistant, Bureau of Educational Research.

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA

1922

BULLETIN NO. 10

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

RELATION OF SIZE OF CLASS TO SCHOOL EFFICIENCY

by

BUREAU OF EDUCATIONAL RESEARCH

Prepared, in part, from a report by P. R. Stevenson, former Assistant, Bureau of Educational Research.

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA

1922

TABLE OF CONTENTS

PREFACE

This investigation was initiated by Mr. B. R. Buckingham, formerly Director of the Bureau of Educational Research. It was planned and executed by him with the assistance of Mr. P. R. Stevenson, a full time assistant in the employ of the Bureau of Educational Research during the school year of 1920-21. The present Director of the Bureau of Educational Research had no connection with the study until late in the summer of 1920-21. The portion of this report which deals with "existing conditions in regard to class size" and "the relation of size of class in high schools to school efficiency" is based upon tabulations made by employees of the Bureau of Educational Research under Mr. Stevenson's direction and included by him in a report submitted to the present Director of the Bureau of Educational Research. The chapter devoted to the relation of the size of class in elementary schools to school efficiency is based upon tabulations made from the original data under the immediate direction of the present Director. The concept of the efficiency ratio and the use of this concept in the interpretation of the data are entirely the work of the present Director of the Bureau of Educational Research. The conclusions are also his own.

Certain limitations of the investigation, which the report discusses in detail, cause the results to have a limited practical significance, but it is thought that the publication is jusified for two reasons. In the first place, the concept of school efficiency and the analysis of the conditions which must be considered in any investigation relating to school efficiency should be helpful to future investigators not only of the question of class size but also of questions of other phases of school procedure. In the second place, the report emphasizes the need for careful planning which will result in the control of all factors involved in the teaching situation. There is also emphasis upon the need for securing normal conditions if the results are to be interpreted with reference to the modification of practise. Such analysis and careful thinking are not only important phases of educational research but they are the foundation upon which both the data collected and the statistical manipulation of them are based.

This investigation was made possible through the cooperation of Superintendent Peter A. Mortenson, of the Chicago Public Schools, and of the school officials in certain other Illinois cities. Not only did they cooperate by permitting the collection of the data, but they actually made substantial contributions to the project by furnishing the test materials. The teachers in the schools concerned made a substantial contribution by scoring the tests and reporting them in a convenient form to the Bureau of Educational Research. The writer is glad to acknowledge the indebtedness of the Bureau of Educational Research to all those who have contributed to the project.

WALTER S. MONROE, *Director.*

May 26, 1922.

RELATION OF SIZE OF CLASS TO SCHOOL EFFICIENCY

CHAPTER I

INTRODUCTION: ANALYSIS OF PROBLEM

The problem. The problem of this investigation is to study the relation of the size of class to school efficiency, or what is the effect upon the efficiency of the school when the size of class is increased or decreased within certain limits.

Definition of terms: Class. In the high school, a class is defined as the number of pupils who are assigned to a single teacher for instruction during a single class period. In the elementary school, unless the instruction has been departmentalized, a class is the number of pupils assigned to a room over which a teacher has charge. For instructional purposes, a teacher in the elementary school may divide a class into two or three groups, but the total number of pupils receiving instruction from her is considered a class, as the term is used in this study.

School efficiency. Educators have borrowed the term "efficiency" from industry and business. In these fields, efficiency is expressed by a fraction whose maximum is 1.00. The numerator of this fraction, or "efficiency ratio," is the output, and the denominator is the input, or educational investment. In education, the output of a school system consists of the changes produced in the pupils, i. e., the controls of conduct that the school engenders. The educational output for a semester or a year is the total of all the changes that have been produced in the pupils during the period due to the influence of the school. The educational investment includes many factors, such as buildings, equipment, textbooks, teachers, supervision, and general administration. Although it is not imperative to do so, it is probably best to think of both the output and the investment as being expressed in terms of the average for one pupil.

We are accustomed to refer to the educational output as the achievements of the pupils. By means of educational tests and other instruments, we measure these achievements in terms of arbitrary units. Units of one type are implied in school marks. Other

units are defined by educational tests. In order to calculate the numerical value of the efficiency ratio, it would be necessary to determine the social value of the output in terms of dollars and cents or in terms of some unit which might be made common to both numerator and denominator. Obviously, we are not prepared to do this. However, we may use the ratio as a definition of school efficiency and inquire into the probable nature of the changes produced in it by the variations of certain factors upon which the educational output and educational investment depend. In making inferences concerning the fluctuations in the value of the efficiency ratio, it is necessary to remember that the production of certain achievements may be of little value when viewed in relation to our educational objectives. The attainment of certain levels of achievement may represent an educational output of considerable value, but advancement to higher levels may produce only slight increases in the value of the total output in this field. For example, the attainment of certain levels of ability in spelling has a distinct and relatively large social value, but advancement beyond these levels is accompanied by rapidly diminishing increments of value. Therefore, one must avoid the assumption that fluctuations in achievements are to be interpreted as having proportional values in terms of social worth.

Factors which affect school efficiency. For a given educational investment per pupil, the value of the efficiency ratio is affected when changes are made in the educational output. Methods of instruction, the plan of school organization, or the procedure of supervision may be modified even when there is no change in the investment. When modifications in the methods of using the investment result in changes in the educational output, there are resulting changes in the value of the efficiency ratio. On the other hand, it is possible that material changes may be made in the educational investment which are not accompanied by corresponding changes in the educational output. When this happens, the value of the efficiency ratio is changed, even though the actual educational output has remained constant.

In many cases, modifications in the method of using the educational investment are accompanied by changes in the magnitude of the educational investment as well as by changes in the educational output. Hence, we may have fluctuations occurring in both the numerator

and the denominator of the efficiency ratio. It is possible that these fluctuations may be connected in such a way that the value of the efficiency ratio remains constant, or it may be that its magnitude will vary. Because we are not able to calculate a numerical value of the efficiency ratio, a careful analysis is required to determine the probable changes in it when variations occur in the numerator and the denominator simultaneously.

The achievements of pupils are materially affected by their general intelligence or capacity to learn. Since individual pupils and also groups of pupils have been shown to exhibit marked individual differences when measured with respect to this trait, it is necessary to make due allowance for differences in general intelligence when comparing different school units with respect to efficiency. In case this is not done an error of interpretation will be made by attributing a higher degree of efficiency to those units which consist of pupils of superior general intelligence.

The effect of varying the size of class. The size of class is one item of the plan of the organization. From this point of view, it may be considered one of the methods of using the educational investment. Consequently, we may expect to find that changes in the size of class produce variations in the achievements of pupils. The size of the class is, also, one of the factors which determines the educational investment. In the elementary school, where a class means the number of pupils assigned to a teacher, the cost of instruction per pupil varies inversely with the size of class.[1] In a high school the size of class does not completely determine the number of student hours of instruction which a teacher gives, but it is a potent factor in this determination. In general, an increase in the size of class in the high school will tend to result in a marked decrease in the educational investment per pupil. Hence, in studying the effect of varying the size of class upon the efficiency of the school, it is necessary for us to inquire into the resulting changes in both the educational output and the educational investment. It is only when we have done this that we are in a position to make inferences concerning the effect of variations in the size of class upon the efficiency of the school.

[1]In making this statement, no account is taken of investments made in supervision, in instruction by special teachers, and in equipment.

The practical importance of a study of the relation of class size to school efficiency. During recent years, school administrators have faced the problem of providing instruction for a rapidly increasing enrollment and, at the same time, of meeting the demands from teachers for increased salaries. In meeting these two demands, there has been a tendency to increase the number of pupils instructed by a teacher in order to keep the total educational expenditure within the income of the school system. In the elementary school, and to a considerable extent in the high school, the number of pupils instructed by a teacher has been increased by increasing the size of classes. It is obvious that pupils in large classes have less opportunity for recitation and, in general, receive less individual attention from the teacher, both within and outside of the class period. Thus, the question has naturally been raised concerning the effect upon the efficiency of a school when the size of the class is increased. In the secondary school added emphasis is given to the question because certain accrediting agencies require that the size of the class not exceed a certain fixed maximum.

Connection between size of class and methods of instruction. It is necessary to bear in mind that, when the number of pupils instructed by a teacher is increased, there is a corresponding increase in the amount of work required of the teacher, unless there are compensating changes in the methods of instruction. For example, according to our present methods, it is customary to require a great deal of written work of pupils studying English composition. A teacher is expected to read with considerable care the compositions submitted by pupils and to provide a systematic procedure for correcting the errors made. Thus, an increase in the number of pupils to be taught increases the work required of the teacher unless the number of compositions required from each pupil is reduced or a different system of handling them is used. Much the same conditions prevail in a number of other subjects in which notebooks or other written work of some sort are customarily required. Similar statements can be made with reference to individual work with pupils outside of the regular class period. It is obvious that there is a limit to the amount of work which may legitimately be required of a teacher. When the optimal teaching load has been reached, any increase in the number of pupils assigned to a teacher should

be expected to be accompanied by compensating changes in the procedure of instruction. When such changes are made in the teaching procedure, they, as well as the size of the class, must be considered in respect to the effect upon the efficiency of the school.

Limitations of the present investigation. In this investigation of the effect of varying the size of class upon school efficiency, it was intended that all other factors which affect either the achievements of pupils or the educational investment should be kept constant, i. e., they should be the same for classes of different sizes. No information is available to show the extent to which this intention was realized except in the case of the quality of the pupil material. Group intelligence tests were used to secure equivalence of capacity to learn in the two groups of pupils. The possibility of a lack of equivalence of such factors as home environment, nationality, attitude toward school work, previous school experience with respect to size of class, time of day (applies only to high school), etc., makes it necessary to exercise due caution in interpreting the results of the investigation.

So far as the present writer is aware, the instruction that prevailed in the groups concerned in this investigation involved no unusual features, and the same methods of instruction were followed in the two types of classes. It is possible that the highest degree of efficiency for classes of a given size would be attained if the methods of instruction were selected with particular reference to this size of class. It would not be surprising to find that methods of instruction which were most effective in small classes would be considerably less effective with large classes, and that methods well suited to the handling of large classes would not give the best results when used with small classes. This possibility was not considered in this investigation. Hence, the findings should not be accepted as final. There is still need for an investigation in which methods of instruction are adapted to size of class.

Finally, it must be remembered that the problem of the size of class is not entirely a problem of the efficiency of the school. It is also a problem of the teacher. It is not humane and it is not socially profitable to assign teaching loads so heavy that teachers become overworked. The problem of the teacher was not considered in this investigation.

CHAPTER II

EXISTING CONDITIONS IN REGARD TO CLASS SIZE

In order that the significance of the two investigations to be reported in the following chapters may be more fully appreciated, certain facts are presented, concerning the size of class under present school conditions in Illinois.

Size of class in elementary schools in Illinois outside of Chicago. Data with reference to the size of class in the elementary schools of the state, outside of Chicago, were secured by sending a questionnaire to the superintendent of public schools in all cities and towns listed in the Illinois School Directory for 1920-21 as having six or more elementary teachers. The questionnaire asked for the number of elementary teachers having classes of the following sizes: less than 20, 20 to 29, 30 to 39, 40 to 49, and 50 and over. This information was requested for each of the school years of 1918-19, 1919-20, and 1920-21. Complete reports were received from 180 cities and towns. These are summarized in Table I. The total number of classes for which a report was secured varied from 9,422, 1918-19, to 10,403, in 1920-21. The median size of class varied from 41.4 pupils, 1918-19, to 43.3 pupils, in 1919-20. In 1920-21, the size of class was slightly less than that for the preceding year. This table also shows that, in the cities reporting, slightly more than 3 percent of all classes contained 50 or more pupils. On the other hand, between 7 and 8 percent of the classes had less than 30 pupils. The greatest change in regard to the size of class during this period was in the marked decrease in the number of classes having between 30 and 39 pupils. During the school year, 1918-19, the size of slightly more than one-third of the classes fell within these limits. During the other two school years, covered by this study, less than one-fifth of the classes came within these limits of size. This change is offset by a corresponding increase in the number of classes having between 40 and 49 pupils. Although the span of years covered by this table is insufficient to justify conclusions with reference to the trend of the size of class, Table I suggests that there is a tendency, in

TABLE I. SIZE OF CLASSES IN THE ELEMENTARY SCHOOLS OF 180 ILLINOIS CITIES FOR 1918-21.

Number in Classes	1918–19		1919–20		1920–21	
	Number of Classes	Percent of Classes	Number of Classes	Percent of Classes	Number of Classes	Percent of Classes
50.....	313	3.3	342	3.5	350	3.4
40–49.....	5134	54.5	6865	70.0	7161	68.8
30–39.....	3264	34.6	1845	18.8	2050	19.7
20–29.....	598	6.3	646	6.6	719	6.9
Less than 20	113	1.3	110	1.1	123	1.2
Total.....	9422	100.0	9808	100.0	10403	100.00
Median.....		41.4		43.4		43.2

the elementary schools of Illinois outside of Chicago, to assign from 40 to 50 pupils to a teacher.

Size of class in Chicago public schools, October, 1920. Because the two investigations to be reported later were carried on, for the most part, in certain schools in Chicago, it was thought desirable to compile separately the facts relating to class size in Chicago. The information was taken from the records of the superintendent of schools. One-half of the elementary schools were selected at random, but all of the high schools were included. Information, with reference to the size of class in each grade, is summarized in Table II. Below the ninth grade, the classes are noticeably larger than in the high school. The median size of class is approximately 46 pupils. In the high school, a greater degree of variability is shown in the size of class. Approximately 14 per cent of the classes have fewer than 20 pupils. The median size of class for the high school is slightly over 30 pupils. It will be noted that the classes for the first year are larger than those for the following years. Of the 183 classes, 178 reported as having from 50 to 54 pupils, are classes in physical training. A number of these have more than 54 pupils.

The size of high school classes is given by subjects in Table III. The median size of class for the different subjects ranges from 20.0,

TABLE II. SIZE OF CLASSES IN CHICAGO SCHOOLS, OCTOBER, 1920.

Grades	Less than 10	Number in class									55 and over	Total	Median
		10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54			
1			3	1	6	20	40	120	248	87	36	561	46.8
2			1	2	1	6	11	102	247	38	1	409	46.7
3			1	1	1	3	18	96	211	48	3	382	46.7
4				1	2	3	15	90	207	39		257	46.6
5			1			2	4	79	215	48	1	350	47.1
6			3	3	1	2	12	103	183	27	4	338	46.2
7					2	2	15	92	180	32	3	326	46.4
8				1		4	54	79	117	15	48	318	44.8
Total Elementary Schools			9	9	13	42	169	761	1618	334	96	3041	46.6
9	36	108	224	556	613	850	683	254	98	92		3474	32.3
10	24	84	202	342	378	401	283	136	17	69		1936	29.3
11	14	60	90	178	148	189	119	31	10	19		858	27.8
12	15	27	64	67	81	93	77	29	2	3		458	28.4
Total H. S.	89	279	580	1143	1220	1533	1162	450	87	183		6726	30.2

for German, up to 34.5, for arithmetic. If physical education is included, the maximum median class is 43.8. However, the most significant aspect of the table is the wide variation in the size of class for a given subject. With few exceptions, there are, in each subject, classes having 10 or less pupils and also classes having more than 45 pupils. The exceptions are shop work, in which the number of pupils is probably limited by the equipment, office practise, of which there are only 37 classes in the entire school system, chemistry, home economics, botany, zoology and agriculture, in which equipment again probably limits the size of class. In German there are no classes having more than 34 pupils, but there are only 8 classes in the entire system.

Opinions of city superintendents in regard to the best size of class. A questionnaire was sent to the city superintendents of public schools in all cities in the United States having a population

TABLE III. SIZE OF CLASSES BY SUBJECTS IN CHICAGO HIGH SCHOOLS, OCTOBER, 1920.

Subject	Less than 10	Number of pupils in class									Total	Median
		10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54		
Botany, Zoology and Agriculture	6	12	24	46	49	51	22	5			215	27.0
French	6	14	33	51	35	38	15	8	2	1	203	24.7
Home Economics	7	27	42	57	54	18	11	3			219	22.9
Geography	1		17	17	19	14	19	8	1		96	28.4
Typewriting	1	1	12	36	50	70	74	29	11		284	33.0
Art	7	22	35	44	44	59	28	11	2		252	27.0
Physiology			1	13	17	36	44	12	1	1	125	34.4
German	2	1	1		3	1					8	20.0
Chemistry	2	5	15	21	19	39	3	1			105	27.5
Office Practise	3	4	2	11	4	6	7				37	24.3
Spanish	2	15	18	28	66	59	41	16	2		247	29.6
Physical Education	5	3	19	21	42	61	40	47	39	178	455	43.8
Latin	3	15	20	31	46	52	56	21	1		245	30.8
Mechanical drawing	22	29	63	136	97	67	31	9	1		455	24.2
Algebra	1	2	19	36	73	133	115	27	2		408	32.7
Geometry	1	9	16	36	62	79	51	13	3		270	30.6
Arithmetic			5	18	25	43	57	24	1		173	34.5
Bookkeeping	2	15	14	23	28	27	30	17	1		157	29.4
Stenography	1	9	23	55	53	72	65	17	4		299	30.6
Physics	1	2	12	29	18	40	22	5	1		130	30.4
Shop	8	66	94	228	101	16	3				516	22.0
History	4	13	27	49	37	92	86	63	3	1	375	33.1
English	3	15	60	143	228	352	278	85	10	2	1176	32.0
General Science	1		8	14	50	108	65	28	2		276	33.0

TABLE IV. IDEAL SIZE OF CLASSES AS INDICATED BY 270 SUPERINTENDENTS IN CITIES OF 25,000 OR MORE POPULATION.

Score	Grades			
	1, 2, 3,	4, 5, 6,	7, 8, 9,	10, 11, 12
55–59	1			
50–54		1	1	
45–49	2	2		1
40–44	16	18	15	2
35–39	53	73	35	4
30–34	95	95	104	33
25–29	52	42	58	95
20–24	20	21	48	95
0–19	30	18	9	16
Total.....	269	270	270	246
Median....	31.7	32.8	31.0	25.6

of 25,000 or more, as shown by the directory issued by the Bureau of Education. The questionnaire asked the superintendents to indicate the ideal size of class in their opinion for grades one to three, four to six, seven to nine, and ten to twelve. Replies from 270 cities are summarized in Table IV. One of the most interesting things about this table is the wide range of opinion which it indicates. A considerable number of superintendents would have fewer than 20 pupils in each class. Other superintendents appear to consider classes of 40 or more ideal. One superintendent indicates that he would be satisfied with classes of 55 pupils in the primary grades. The ideal median size of class below the high school ranges from 31.0, for the junior high school, to 32.8, for the intermediate grades. In the senior high school, the ideal median of class is 25.6 pupils.

The problem of class size. Although Table IV is based upon replies from superintendents distributed over the United States, and the preceding tables refer to existing conditions in Illinois, we are probably justified in pointing out that a marked difference exists between theory and practise. The prevailing practise in the elementary schools of the state outside of Chicago centers around classes

having from 40 to 45 pupils. The median ideal size of class is only slightly above 30. Thus, it appears that the practical problem in which school superintendents are interested relates to a determination of the relative efficiency of classes enrolling from 25 to 35 pupils as compared with classes enrolling from 40 to 45 pupils.

For the high school, we have no data relating to the size of class in Illinois except for the city of Chicago. The median ideal size is approximately 25. The median actual size is approximately 30. It, therefore, appears that the practical problem in the high school relates to the relative efficiency of classes enrolling from 20 to 30 pupils as compared to those enrolling from 25 to 35.

CHAPTER III

RELATION OF SIZE OF CLASS IN ELEMENTARY SCHOOL TO SCHOOL EFFICIENCY

General plan of the study in the elementary school. In order to study the effect of variations in the size of class upon the achievements of pupils, it is necessary to hold constant or to measure the other factors which affect their achievements. It was planned to keep the teacher constant by having both a large class and a small class taught by the same teacher. Since the plan of organization in the elementary school makes it impossible for the same teacher to instruct two classes at the same time, it was necessary to have a teacher instruct the two types of classes during two consecutive semesters. It was arranged to have some of the teachers instruct a large class during the first semester and a small class during the second semester. Other teachers instructed a small class during the first semester and a large class during the second semester.[1]

In order to keep the pupil material as nearly constant as possible, "one hundred percent promotion" was secured at the end of the first semester in all of the experimental groups. When a teacher instructed a large class during the first semester a number of pupils were sent to another teacher at the beginning of the second semester. The pupils remaining formed a small class. In doing this, an effort was made to select pupils so that those remaining would form a small class having approximately the same average mental age and the same variability of this trait. When a teacher instructed a small class during the first semester, pupils were added at the beginning of the second semester, but care was exercised to have these pupils such that the average mental age of the class would not be materially affected.

This investigation, which began in October, 1920, was confined to classes in grades II, V, and VII. Some of the experimental groups were organized in the B sections of these grades and the others in the A sections. At the beginning of the second semester the B sec-

[1]This investigation was carried on in five elementary schools in Chicago: Washington, Cleveland, Lowell, Farragut, and Hibbard.

tions became A sections of the same grade and the A sections became B sections of the next higher grade. We shall, however, refer to the grades simply as II, V, and VII. Data for only those pupils who attended both the large class and the small class and who took all of the tests given are included in the following tabulations.

The size of the experimental classes. In the second grade there were eleven experimental classes, three small the first semester and large the second, and eight of the opposite type. If one class, enrolling only 18 pupils when considered a large class, is excluded, the small classes range from 33 to 44 and the large from 45 to 54. The differences in the size of the paired groups range from 4 to 13, the average being approximately eleven pupils. In the fifth grade there were thirteen experimental classes, three small the first semester and large the second, and ten of the opposite type. The small classes range in size from 33 to 45 and the large from 42 to 52. The differences in the size of the paired groups range from 4 to 14, the average difference being approximately 9. In the seventh grade there were only five experimental classes, three of one type and two of the other. The small classes ranged from 35 to 44 and the large from 42 to 49. The average difference in size was approximately 7. It should be noted that in both the fifth and seventh grades there is some overlapping in the size of the two types of classes. Some "large classes" are smaller than certain "small classes."

Data collected. In the second grade, the Dearborn Group Intelligence Test and Pressey Primer Scale were given at the beginning of the experiment. In this grade, achievement was measured by giving the Indiana Scale of Attainment, No. 1. Form 1 was given in October, Form 2 in January, and Form 1 was used again at the end of the year. In grades V and VII, the Illinois General Intelligence Scale was used. The achievements of the pupils were measured in arithmetic, silent reading, language, and spelling. In arithmetic and reading, the measurements were secured by means of the tests included in the Illinois Examination. Form 1 was used for the first and third testings and Form 2 for the second. In language, Charters' Diagnostic Language Test for pronouns was used. Form 1 was given in October and in May. Form 2 was used for the January testing. In spelling, 20 words were selected from columns N and R of the Buckingham Extension of the Ayres' Spelling Scale.

For the first and third testings, the words were selected by beginning at the bottom of these columns and choosing alternate words. The words for the second testing were taken from these columns, beginning with the next to the last word and taking alternate ones.

Administration of tests and collection of data. All the tests were administered and scored by the teacher. As a preparation for this work, the teachers were called together and given definite instructions concerning the nature of the tests and the plan of administration. In this connection, the tests were administered to the teachers in order to illustrate to them the procedure to be used with the pupils. All of the tests are highly objective with reference to the scoring, and samplings of the test papers failed to reveal any large errors in this work. The teachers reported the data for each pupil on an individual record card. This card contained spaces for each score and each test as well as for data with reference to the size of class in which the pupil was taught during each semester. The dates of testing were approximately as follow: October 20th, February 20th, and May 20th.

Method of summarizing the data. The data summarized were limited to the scores of only those pupils who were members of both a large class and a small class, and who were present at all three testing periods. The scores of all pupils in a grade (including both A and B sections), who had been taught in the same type of class, were assembled for each of the three testings. For example, the October arithmetic scores for all fifth grade pupils who were taught in large classes during the first semester and in small classes during the second semester were assembled in one distribution. Another distribution was made for the January scores and a third one for the May scores. Three corresponding distributions were made for the arithmetic scores of the pupils who were taught in small classes during the first semester and in large classes during the second. Thus, there were obtained six distributions for each achievement score. One measure of the gain in achievement made by a group of pupils during the first semester was found by subtracting the *average* of the October scores from the *average* of the January scores. Another measure of the gain was found by subtracting the *median* October score from the *median* January score. In a similar manner, the gains for the second semester were obtained by sub-

tracting the average and the median January scores from the corresponding May scores.

In calculating these gains, no account was taken of the possible non-equivalence of the different forms of the tests used. In fact, no accurate information concerning the equivalence of the duplicate forms is available, except for the tests in reading and arithmetic. The duplicate forms of these two tests have been shown to be approximately equal.[2] Since Form 1 was used twice, and the average and the median scores calculated from it were used both as subtrahends and minuends, any non-equivalence of the forms will not affect the comparisons of gains made in the following table.

The scores of the different tests are expressed in terms of different units. Thus, before any combination of the results from the different tests can be made, it is necessary to express the gains in terms of a common unit. The usual assumption in such cases is that the standard deviation of the distribution of scores represents the same increment of ability for one test and in one grade as in another. On the basis of this assumption, the stardard deviation was calculated for six of the different distributions of scores for each test in a given grade, and the average of these six standard deviations was used as a divisor to reduce the gains to a basis of a common unit. For example, during the first semester the fifth grade pupils taught in large classes in arithmetic made a gain of 12.0 points.[3] During the second semester they made a gain of 7.2 points. The gains for the pupils taught in small classes in arithmetic were 6.5 and 5.65. The average standard deviation of the six distributions of arithmetic scores is 17.441. Dividing these gains by this average standard deviation, we secure as quotients the entries (.68, .41, .36, and .32) to be found in Table V.

In calculating the average gains for the two types of classes, the simple average of the two gains has been used rather than the weighted average, although the two groups are not even approximately equivalent in size. Since our purpose in taking this average is to eliminate any differences in the course of study or in the edu-

[2]Monroe, Walter S. "The Illinois Examination." University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin, No. 6. Urbana: University of Illinois, 1921.

[3]These gains were calculated from the average of the scores.

TABLE V. THE GAINS MADE BY PUPILS IN LARGE CLASSES AND IN SMALL CLASSES. CLASSES OF ALL SIZES INCLUDED

	No. of pupils	Average size of class		Arithmetic Gains		Dif.	Reading Comprehension Gains		Dif.	Reading Rate Gains		Dif.	Language Gains		Dif.	Spelling Gains		Dif.	Totals		Dif.
		Large	Small	Large	Small		Large	Small		Large	Small		Large	Small		Large	Small		Large	Small	
GRADE II.																					
Average Scores:																					
Large 1st S.....	244	44.8	33.1	1.89	.59		1.17	.59		1.27	.38					1.65	.88		5.98	2.44	
Small 1st S.....	109	48.0	38.0	.30	1.48		.53	1.09		.51	1.00					−.13	1.83		1.21	5.40	
Average.....				1.09	1.04	.05	.85	.84	.01	.89	.69	.20				.76	1.36	−.60	3.59	3.93	−.34
Median Scores:																					
Large 1st S.....	244	44.8	33.1	2.25	.70		1.28	.80		1.09	.60					2.06	.95		6.68	3.05	
Small 1st S.....	109	48.0	38.0	.33	1.94		.80	1.05		.53	.97					.50	1.64		2.16	5.60	
Average.....				1.29	1.32	−.03	1.04	.93	.11	.81	.78	.03				1.28	1.29	−.01	4.42	4.32	.10
GRADE V																					
Average Scores:																					
Large 1st S.....	325	45.8	35.7	.68	.41		.20	.88		.09	.72		.89	1.22		.02	.62		1.88	3.85	
Small 1st S.....	132	51.0	44.0	.36	.32		.55	.38		.76	.28		.93	.96		.38	.29		2.98	2.23	
Average.....				.52	.37	.15	.38	.63	−.25	.43	.50	−.07	.91	1.08	−.17	.20	.46	−.26	2.44	3.04	−.60
Median Scores:																					
Large 1st S.....	325	45.8	35.7	.78	.36		.32	1.12		.08	1.14		.79	1.77		.01	.67		1.98	5.06	
Small 1st S.....	132	51.0	44.0	.21	.43		.75	.29		.93	−.03		.98	.99		.43	.26		3.30	1.94	
Average.....				.50	.40	.10	.54	.71	−.17	.51	.56	−.05	.89	1.38	−.49	.22	.46	−.24	2.66	3.51	−.85
GRADE VII.																					
Average Scores:																					
Large 1st S.....	104	45.0	36.0	.85	1.29		.69	.72		.74	.45		.57	1.26		.92	.21		3.77	3.93	
Small 1st S.....	73	47.0	42.0	.53	.47		.42	.19		.64	−.61		.62	1.00		.15	.44		2.36	1.49	
Average.....				.69	.88	−.19	.56	.46	.10	.69	−.08	.77	.59	1.13	−.54	.54	.33	.21	3.07	2.72	.35
Median Scores:																					
Large 1st S.....	104	45.0	36.0	.92	1.28		.76	.86		.67	1.01		.50	1.34		.98	.47		3.83	4.96	
Small 1st S.....	73	47.0	42.0	.65	.66		.45	.21		.56	−.45		1.09	.99		.24	.64		2.99	2.05	
Average.....				.79	.97	−.18	.61	.54	.07	.62	.28	.34	.79	1.17	−.38	.61	.56	.05	3.42	3.52	−.10
Totals of averages of Average Scores...				2.30	2.29	.01	1.79	.193	−.14	2.01	1.11	.90	1.50	2.21	−.71	1.50	2.15	−.65	9.10	9.69	−.59
Totals of averages of Median Scores....				2.58	2.69	−.11	2.19	2.18	.01	1.94	1.62	.32	1.68	2.55	−.87	2.11	2.31	−.20	10.50	11.35	−.85

cational opportunities offered in the two semesters and also any practise effect due to acquaintance with the tests, it seemed unwise to weight the averages on the basis of the number of pupils in the two groups. To have used the weighted averages in this case would have resulted in giving greater weight to the gains made by one group of pupils simply because this group happened to be larger.

Achievements of the two groups approximately equal. Table V[4] summarizes the data with reference to the gains made by the two groups. The column headed "Number of pupils" gives the number of pupils whose records were used in the tabulations. (A pupil's record was discarded if he was not a member of both the small class and the large class and if he did not take all tests.) The average size of class is computed from the total enrollment. The computation of the gains has just been explained. In interpreting the table, attention should be focused upon the differences. A positive difference means that the large class is superior in achievement, and a negative difference that the small class is superior. At the bottom of each difference column the differences, calculated from the averages and also from the medians, are summarized. This summary is a total and not an average. To find the average it is necessary to divide by 3. All of the totals of the differences fall between +1.00 and —1.00. Six of the 10 differences are negative, and only in the case of reading rate are the difference between the averages and the difference between the medians both positive. In the last three columns of the table, we have the totals and not averages. To find the average, it would be necessary to divide by 4 in the second grade and by 5 in each of the other two grades. Here, again, the negative differences predominate, although none of them are very large. The last two entries in the last column are essentially grand totals and may be considered to summarize the entire table. To find the average difference, each of these numbers should be divided by 14. The quotients obtained would be —.04 and —.06. Thus, in general, this table indicates that there is little if any superiority in the achievements of pupils in the small classes over those of pupils in the larger classes.

---

[4]The entries in the column headed "Reading Rate" in the second grade are based upon the Pressey Word Recognition Test.

An examination of Table V reveals the fact that, in general, the gains between the first and second testings are much larger than those between the second and third testings. This condition emphasizes the necessity for equalizing the effect of acquaintance with the test and practise effect. If all of the experimental groups had been taught as large classes the first semester and small classes the second, the gains for the large classes would greatly exceed the gains for the small classes; but this would be due primarily to the effect of acquaintance with the test and the practise effect.

When Table V is examined with reference to the conditions in the different grades we find that the gains are relatively greater for the small classes in the fifth grade than in either the second or the seventh grade. However, the number of pupils is so small for the two groups in the seventh grade and the groups differ so little in size that only slight significance can be attached to the results. Even in the second and fifth grades it is unfortunate that the experimental groups are not more nearly equal in size. It is possible that if the experiment had included a larger number of classes which were small the first semester and large the second different results might have been obtained.

When the gains for the different subjects are examined we find that in language the gains for the small classes are consistently greater than the gains for the large classes. This is not true for any other subject. Although in spelling the total of the gains is distinctly negative, in both arithmetic and reading comprehension the total when computed by one method is negative and in the other case is approximately 0. Reading rate in the seventh grade is the only case in which the large class is distinctly superior in achievement.

Conclusion: the relation of the size of class to school efficiency. Since Table V indicates that, on the whole, there is little difference between the achievements of the pupils when taught in large classes and their achievements when taught in small classes, one might infer that the efficiency of a school would be materially increased by the formation of large classes, because the educational output would be approximately the same and the educational investment would be materially decreased. However, it is doubtful that the present investigation justifies such a conclusion. In the first place, it is obvious that only certain achievements of pupils have

been measured. Even in the fields of the four subjects in which tests were given we are not justified in claiming that all achievements of the pupils were measured. The arithmetic tests used were confined to the operations and to only certain types of examples within this division of arithmetic. In silent reading, the test used is very limited in scope. Similarly, the tests in language and spelling possess very definite limitations with respect to scope. There is some justification for assuming that the measurements made may be considered indices of the total achievements of the pupils not only in the fields of the four subjects in which the tests were given but also in the field of instruction in the grades concerned. However, the thesis that the measures of achievement secured in this investigation are indices of the total achievement is largely an assumption, and in interpreting the results it is necessary to recognize this fact. It is possible that, if other tests had been used or if the achievements of the pupils had been more completely measured by including tests in other subjects, the results might have been different.

In the second place, it must be remembered that the size of the "small classes" was not less than 33 (with one exception), and in a few cases the enrollment was as much as 44 or 45. The large classes ranged in size from 42 to 54. The average difference between the pairs of experimental groups ranged from 7 in the seventh grade to 11 in the second grade. These conditions with reference to the size of the experimental groups constitute a very significant limitation of the investigation. One is not warranted in making inferences from the facts of Table V with reference to the relative efficiency of classes of 20 to 25 pupils as compared with classes of 35 to 45 pupils. No application should be made except within the limits of size defined by the experimental groups.

CHAPTER IV

RELATION OF SIZE OF CLASS IN HIGH SCHOOL TO SCHOOL EFFICIENCY

General plan of the study in the high school. In the investigation in the high school it was arranged to have both a large class and a small class in a given subject instructed by the same teacher during the same semester. This made it necessary to restrict the experiment to teachers who were instructing two or more sections of the same subject. When a teacher was instructing two sections, pupils were shifted on the basis of their intelligence scores so that the average quality of pupil material was approximately the same in the two sections. Thus, both classes would include some bright, some medium, and some dull pupils. When a teacher had four sections of the same subject, the pupils were shifted so that a large class and a small class would be obtained, consisting of relatively bright pupils. The less capable pupils were also divided into two classes, one large and one small.

In establishing the two types of class, there was considerable variation in the size of both the large classes and the small classes. The small classes varied in size from 12 to 26 pupils. The large classes varied in size from 23 to 45 pupils. The average size of the large classes was 36.5 and that of the small classes, 20.8. The differences in the size of the paired classes ranged from 6 to 26.

Source of data. This study was carried on in four large high schools in Chicago and in the high schools of three other Illinois cities.[1] During the first semester of 1920-21, the experiment was carried on in beginning tenth grade classes. During the second semester of that year the study was confined to classes completing the first year of high school work. In the following tables no distinction is made between classes for the two grades. Records were secured for 67 pairs of classes, enrolling 3,821 pupils. The project

[1]The high schools in Chicago were Lane Technical, Tilden Technical, Harrison Technical, and Hyde Park. The three other Illinois cities were Macomb, Shelbyville, and West Aurora.

was begun by giving the Terman Group Test of Mental Ability to all pupils concerned, about October 15, 1920. Some pupils in the high schools outside of Chicago were given the Illinois General Intelligence Scale or the Chicago Group Intelligence Test. As soon as the results of the testing could be assembled, the large classes and small classes were arbitrarily formed, following the method indicated above.

The educational output, or the achievements of the pupils, was measured by requiring each teacher to give the same final examination to both types of classes. A check upon this measurement of achievement was secured by using the "term grades" of the pupils. It is generally recognized that "term grades," as well as examinations set by teachers, are highly subjective. However, in this case the same teacher administered the same examination to both the small class and the large class. The same teacher also gave the "term grades." Thus, there is in no place a comparison between either "term grades" or "examination grades" given by different teachers. This tends to eliminate the subjective factor of these measures. In addition, it may be noted that we are concerned with the average "grades" of relatively large groups of pupils and not with the "grades" of individual pupils.

A limitation. The plan of carrying on the experiment implies the assumption that the achievements of the pupils in the large classes were equal to the achievements of the pupils in the small classes at the beginning of the experimental period. There was no attempt to measure the achievements of pupils in the subjects concerned at the beginning of the experiment. In the first semester, the two types of classes were not organized until after the high schools had been in session several weeks. In the case of the classes used during the second semester, the pupils had received an entire semester of instruction in regular classes. It is true that the sections were formed so that the average general intelligence of the paired classes was approximately equal, but this probably does not justify the assumption of equivalent achievements.

Details of administration. The intelligence tests were administered and scored by the teachers. As preparation for this work, the teachers were called together and given definite instructions concerning the nature of the tests. The tests were also administered

to them. They were then required to score their own papers. All the tests used were highly objective with reference to scoring, and a sampling of the test papers of the pupils failed to reveal any large errors in this work. Since, in the use of the intelligence test scores, comparisons are always made between the pupils or groups of pupils under the same teacher, variations in the administration of the tests, due to differences between teachers, would not be significant.

At the close of each semester the teachers were asked to report both the "final grade" and the "examination grade" for each student.[2] In most cases, the "examination grades" were reported in terms of percents. The "final grades" were generally reported in terms of letters, as follow:

S=Superior
E=Excellent
G=Good
F=Fair
D=Failure

For the purpose of combining "grades," these letters were assumed to represent the following numerical ranges on a scale of 100 percent:

S is equivalent to 95 and over
E is equivalent to 85 to 94
G is equivalent to 75 to 84
F is equivalent to 65 to 74
D is equivalent to 55 to 64[3]

Plan of summarizing data. The data collected were summarized to show the differences, if any, which existed between the final achievements of pupils in the large classes and of pupils in the small classes. Two methods of summarizing were employed. In the first, the achievements of all pupils were considered. According to the second method, the records considered were limited to those of pupils in the large classes who could be paired with pupils having identical scores on the intelligence test in the corresponding small

[2]The following tables, in which the data for high school classes are summarized, indicate that the "examination grades" were not received from certain classes.

[3]The midpoints of these intervals were presumably used as the numerical equivalents of the grades expressed in terms of letter. However, Mr. Stevenson's report yields no information on this point.

classes. For both of these methods two tabulations have been made. The first includes all classes, and the second only those pairs of classes in which the large class was at least twice the size of the small class.

Differences in achievements when all pupils are considered. Table VI illustrates the first method of summarizing the data for 22 pairs of English classes. In the second and third columns of this table the enrollment in the large classes and in the small classes is given. The quantities recorded in the three columns headed "Difference" are found by subtracting the quantities in the two columns immediately preceding the difference column. The number for the small class is, in every case, taken from that for the large. Therefore, a positive difference means that the large class is superior in the trait concerned, and a negative difference, that the small class is superior. The line at the bottom of the table gives the average for each column. These averages may be taken as summarizing the data collected from these 22 pairs of English classes, although the items combined are not entirely comparable. For example, different general intelligence tests were used in different classes. The present writer has not been able to ascertain the particular intelligence test given to any pair of these 22 pairs of classes. It is, however, difficult to explain the extreme differences between the average intelligence scores of classes 10 and 11 on any basis other than the use of different tests in these two pairs of classes. Furthermore, it is not unlikely that different passing marks are in use in the different schools in which the pairs of classes were taught. If this is the case, in the case of different pairs of classes, both the average term grades and the average examination grades are on different scales.

It should be noted that, although an effort was made to organize a large class and a small class so that the average intelligence scores would be approximately the same for the two classes, this was not always accomplished. Because of conflicts or other conditions that could not be disregarded, it was not always possible to shift pupils from one section to the other so as to set up the desired class organization.

Table VII summarizes the averages for the classes in the different subjects. In interpreting this table, it is necessary to bear in mind that, with the exception of English and algebra, the number of

TABLE VI. COMPARISON OF AVERAGE INTELLIGENCE SCORES, AVERAGE TERM GRADES, AND EXAMINATION GRADES OF 22 LARGE AND 22 SMALL ENGLISH CLASSES.

Teacher	Size		Average intelligence score		Difference	Average term grade		Difference	Average examination grade		Difference
	Large class	Small class	Large class	Small class		Large class	Small class		Large class	Small class	
1	36	12	58.7	57.2	1.5	85.8	77.5	8.3	85.3	77.5	7.8
2	37	17	95.2	99.7	−4.5	69.4	71.2	−1.8			
3	34	17	54.0	50.7	3.3	76.8	79.4	−2.6			
4	40	18	111.5	114.2	−2.7	82.1	81.2	0.9	77.8	85.1	−7.3
5	44	18	92.3	88.4	3.9	72.8	72.2	0.6	74.1	73.6	0.5
6	38	19	98.1	100.2	−2.1	74.2	80.6	−6.4	71.6	80.0	−8.4
7	35	19	138.7	137.1	1.6	77.1	81.6	−4.5	79.8	83.2	−3.4
8	32	19	94.9	87.8	7.1	72.4	75.6	−3.2	69.2	74.1	−4.9
9	41	20	77.4	72.7	4.7	75.8	77.0	−1.2	75.8	75.3	0.5
10	38	20	90.1	89.7	0.4	75.8	77.0	−1.2	80.0	78.0	2.0
11	35	21	55.8	55.7	0.1	70.9	73.0	−2.1	71.2	75.8	−4.6
12	43	21	136.1	142.0	−5.9	79.8	82.7	−2.9	78.3	81.1	−2.8
13	41	21	116.1	110.0	6.1	75.0	73.3	1.7	75.3	72.4	2.9
14	38	21	120.3	122.1	−1.8	87.5	86.9	0.6	88.3	86.8	1.5
15	36	21	97.4	98.4	−1.0	75.2	75.9	−0.7			
16	39	21	95.5	95.6	−0.1	76.4	82.9	−6.5	77.4	83.0	−5.6
17	41	22	131.2	119.9	11.3	76.8	77.8	−1.0	78.4	74.6	3.8
18	40	22	101.0	89.6	11.4	80.4	80.0	0.4	76.7	80.2	−3.5
19	32	24	127.1	143.7	−16.6	76.7	75.8	0.9	81.5	81.0	0.5
20	37	24	130.3	127.1	3.2	79.7	78.6	1.1	80.7	82.9	−2.2
21	38	26	105.7	99.3	6.4	78.3	82.3	−4.0	80.3	83.8	−3.5
22	38	26	130.4	135.2	−4.8	70.5	76.5	−6.0	52.2	69.0	−16.8
Total.....	833	449	2257.8	2236.0	21.8	1689.4	1719.0	29.6	1453.9	1497.4	−43.5
Average....	37.9	20.4	102.6	101.6	1.0	76.8	78.1	−1.3	76.5	78.8	−2.3

pairs of classes is so small that we probably should not consider the result representative. Certainly, little if any significance can be attached to the results for Latin, history, and French. The average for the 67 pairs of classes is the weighted average, so that one pair of classes does not influence this average any more than any other pair.

Table VIII presents a summary for those pairs of classes in which the large class is at least twice the size of the small class. By doing this, we are able to examine the achievements of pupils in pairs of classes where the difference in size is marked. The number of classes in any one school subject is so small, with the exception of English, that we are probably not justified in drawing any conclusions for the separate subjects.

Differences in the achievements of paired pupils. Table IX is similar to Table VII, the only difference being that it is based upon the records of only those pupils in the large classes who could be paired with pupils in the corresponding small classes, having the same scores on the general intelligence test. Since the pupils were paired on the basis of their intelligence scores, the average general intelligence of those taken from the large classes would be identical with the average general intelligence of those taken from the corresponding small classes. Hence, the average general intelligence scores are omitted. Table X is similar to Table VIII.

Interpretation of the tables. When all 67 pairs of classes are considered, the average of the general intelligence scores for the large classes is almost identical with that for the small classes, the difference being only one-tenth of a unit. This unit corresponds approximately to one month of mental age. We may, therefore, consider the pupils in the large classes equal in general intelligence to the pupils in the small classes. Both the average "term grade" and the average "examination grade" are slightly larger for the small classes.

The question of the significance of the difference of two averages is involved here. Both examination grades and final grades are known to be highly subjective and to involve a large error which is a combination of a constant error and a variable error. Variable errors tend to offset each other in an average because some of them are negative and some positive. On the other hand, constant errors

TABLE VII. COMPARISON OF AVERAGE INTELLIGENCE SCORES, AVERAGE TERM GRADES, AND AVERAGE EXAMINATION GRADES OF THE DIFFERENT SUBJECTS FOR 67 LARGE AND 67 SMALL CLASSES.

Subject	No. of pairs of classes	Average size		Average intelligence score		Difference	Average term grade		Difference	Average examination grade		Difference
		Large class	Small class	Large class	Small class		Large class	Small class		Large class	Small class	
English	22	37.9	20.4	102.6	101.6	1.0	76.8	78.1	−1.3	76.5	78.8	−2.3
Algebra	18	35.7	20.3	99.6	102.2	−2.6	74.3	76.5	−2.2	76.3	77.2	−0.9
General Science	6	35.7	21.5	110.5	109.6	0.9	75.3	77.2	−1.9	74.8	76.1	−1.3
Botany	4	28.3	18.0	126.3	122.2	4.1	73.8	74.2	−0.4	71.9	72.9	−1.0
Stenography	3	36.7	22.0	83.6	85.9	−2.3	77.6	78.1	−0.5	78.4	80.5	−2.1
Geometry	3	40.7	22.7	128.1	121.1	7.0	75.8	76.6	−0.8	62 9	67.8	−4.9
Arithmetic	3	38.7	22.7	90.2	89.8	0.4	74.7	74.5	0.2	63.3	64.3	−1.0
Typewriting	3	39.3	24.7	83.4	87.9	−4.5	73.2	76.6	−3.4	80.1	83.0	−2.9
French	2	35.0	19.0	110.0	113.2	−3.2	73.7	70.9	2.8	71.2	71.4	−0.2
History	2	33.5	21.0	139.1	137.4	1.7	75.7	74.6	1.1	74.6	74.3	0.3
Latin	1	39.0	25.0	103.7	92.5	11.2	77.3	73.9	3.4	75.4	70.9	4.5
Average		36.5	20.8	104.1	104.0	0.1	75.4	76.7	−1.3	74.8	76.4	−1.6
Total	67											

TABLE VIII. COMPARISON OF AVERAGE INTELLIGENCE SCORES, AVERAGE TERM GRADES, AND AVERAGE EXAMINATION GRADES OF LARGE AND SMALL CLASSES WHERE THE LARGE CLASSES ARE AT LEAST TWICE THE SIZE OF THE SMALL CLASSES.

Subject	No. of pairs of classes	Average size		Average intelligence score		Difference	Average term grade		Difference	Average examination grade		Difference
		Large class	Small class	Large class	Small class		Large class	Small class		Large class	Small class	
English	8	39.1	17.8	90.4	90.6	−0.2	77.1	77.7	−0.6	77.2	78.8	−1.6
Algebra	5	39.6	18.2	96.2	100.4	−4.2	74.6	78.5	−3.9	85.6	86.3	−0.7
General Science	1	42.0	17.0	73.4	78.7	−5.3	70.4	70.0	0.4	70.4	70.0	0.4
Botany	1	33.0	15.0	128.8	117.6	11.2	78.1	71.8	6.3	78.4	75.5	2.9
Stenography	1	36.0	18.0	52.4	55.7	−3.3	73.3	72.9	0.4	71.3	76.5	−5.2
Average		38.9	17.7	91.2	92.4	−1.2	75.7	76.8	−1.1	79.1	80.2	−1.1
Total	16											

TABLE IX. COMPARISON OF AVERAGE TERM GRADES AND AVERAGE EXAMINATION GRADES OF PUPILS PAIRED ACCORDING TO EQUAL INTELLIGENCE SCORES IN 67 LARGE AND 67 SMALL CLASSES.

Subject	Number of pairs of classes	Number of pairs of pupils	Average size		Average term grade of paired pupils		Difference	Average examination grade of paired pupils		Difference
			Large class	Small class	Large class	Small class		Large class	Small class	
English	22	336	37.9	20.4	76.4	78.1	−1.7	76.6	78.8	−2.2
Algebra	18	298	35.7	20.3	74.3	76.3	−2.0	76.4	77.1	−0.7
General Science	6	102	35.7	21.5	76.2	78.7	−2.5	75.5	76.6	−1.1
Botany	4	36	28.3	18.0	73.5	72.9	0.6	68.8	70.5	−1.7
Arithmetic	3	56	38.7	22.7	73.8	74.3	−0.5	62.3	64.7	−2.4
Geometry	3	50	40.7	22.7	76.1	77.0	−0.9	61.9	68.2	−6.3
Stenography	3	44	36.7	22.0	76.3	79.8	−3.5	76.5	81.6	−5.1
Typewriting	3	42	39.3	24.7	73.7	76.2	−2.5	77.9	83.6	−5.7
French	2	25	35.0	19.0	71.2	70.0	1.2	68.9	70.0	−1.1
History	2	31	33.5	21.0	78.1	75.4	2.7	75.1	75.4	−0.3
Latin	1	14	39.0	25.0	73.6	79.3	−5.7	73.6	75.0	−1.4
Average			36.5	20.8	75.2	76.8	−1.6	74.4	76.5	−2.1
Total	67	1034								

TABLE X. COMPARISON OF AVERAGE TERM GRADES AND AVERAGE EXAMINATION GRADES OF PUPILS PAIRED ACCORDING TO EQUAL INTELLIGENCE SCORES IN LARGE AND SMALL CLASSES IN WHICH THE LARGE CLASSES ARE TWICE THE SIZE OF THE SMALL CLASSES.

Subject	Number of pairs of classes	Number of pairs of pupils	Average size		Average term grade of paired pupils		Difference	Average examination grade of paired pupils		Difference
			Large class	Small class	Large class	Small class		Large class	Small class	
English	8	116	39.1	17.8	76.6	77.7	−1.1	76.7	78.5	−1.8
Algebra	5	84	39.6	18.2	73.7	77.5	−3.8	83.6	86.3	−2.7
General Science	1	16	42.0	17.0	71.9	71.9	0.0	71.9	71.9	0.0
Stenography	1	11	36.0	18.0	72.7	72.7	0.0	70.6	77.6	−6.0
Botany	1	11	33.0	15.0	74.6	71.8	2.8	71.8	75.5	−3.7
Average			38.9	17.7	75.0	76.6	−1.6	77.9	80.3	−2.4
Total	16	238								

are not eliminated in averages. The number of pupils included in the 67 pairs of classes is sufficiently large so that the average grades include a variable error which is probably so small as to be negligible. For example, if we assume that the probable variable error of a grade is as much as 10 points, which is probably in excess of the actual probable variable error, the probable variable error of the average for the small classes would be less than three-tenths of one point. In the case of the large class, it would be materially less than in the small class. The constant errors, which are expressed in the tendency of some teachers to give higher grades than others, are probably included in both groups in approximately the same proportion, since the same teacher assigned grades to both a small class and a large class. If this is true, a difference of 1.3 between the average term grades is probably significant, although there is a reasonable doubt. This doubt is materially increased and probably the difference loses its significance when we recall that the achievements of the two groups of students were not measured at the beginning of the experimental period.

When we consider the summaries for the different subjects, we find the differences between the average achievements of the pupils in the two types of classes materially larger in a number of cases than the differences between the averages for the 67 pairs of classes. However, in interpreting these differences it is necessary to remember that the number of pupils is materially less and, hence, a difference must be materially greater in order to be significant. It is perhaps significant that negative differences predominate. This suggests that the average achievements in the small classes are somewhat superior to those in the large classes, but in drawing conclusions from this condition it is necessary to bear in mind the fact that the achievements of the students were not measured at the beginning of the experimental period.

When we turn to Table VIII, which is restricted to those classes in which the large class is at least twice the size of the small class, we find that the pupils in the small classes were slightly superior in general intelligence. If the size of class is a potent factor in determining the achievements of pupils, we should naturally expect to find a greater difference in the achievements of the two types of classes than we found in Table VIII. The fact that the difference

is less, both absolutely and relatively, suggests that the size of class is not a potent factor in determining the achievements of pupils. However, the number of pairs on which Table VIII is based is so small that no great importance should be attached to this observation.

When we examine Table IX, we find that the differences between the averages for the 67 pairs of classes are only slightly larger than those given in Table VII. In general, there are few significant differences to be noted in a comparison of these two tables. One of the most significant is the reversal of the relative achievements of the large class and the small class in Latin. In Table VII, when all pupils were considered, those taught in the large class were shown to be distinctly superior in achievement. When only the paired pupils were considered, those taught in the small class were distinctly superior in achievement. Considering the table as a whole, we are justified in asserting that it tends to corroborate the interpretations suggested for Table VII.[4]

Conclusion: relation of size of class in high school to school efficiency. The tables of this chapter show that at the end of the experimental period the achievements of the students in the two types of classes were approximately equal, and there is a slight indication that those taught in small classes were superior. Since the educational investment can be materially decreased by increasing the size of class in the high school, one might infer that the efficiency of the school would be increased by organizing classes enrolling from 35 to 40 students instead of classes enrolling from 20 to 25. In addition to the fact that there are several uncontrolled factors whose influence is unknown, it is necessary to bear in mind the exact conditions of the experiment. Since the same teachers taught both

[4]In his report, Mr. Stevenson attempted the further analysis of the data by ascertaining the percent of pairs of pupils in which the pupil in the small class received a higher grade than the corresponding pupil in the large class. When all pairs of pupils are considered, he shows that only in 40 percent of the pairs does the pupil in the small class surpass his mate in the large class. However, it is impossible to draw any conclusion from this fact, because we do not know the percent of pairs in which the two pupils received the same mark. Since, in a majority of the cases, the grades were reported in terms of letters and only five marks were recognized, it is reasonable to expect that in a relatively large percent of the cases both of the paired pupils received the same final grade.

a small class and a large class, there was no difference between the total amount of work done by the teachers who handled the large classes and the teachers who handled the small classes. In fact, they were the same teachers. Thus, this experiment failed to set up the conditions of large classes as a general plan of organization of a high school. It did, however, realize the conditions which not infrequently exist in the smaller high schools where it is desirable to have a few large classes assigned to teachers who are given compensating small classes or who have the number of classes reduced accordingly. The results of the experiment, therefore, can be applied only to those situations in which the teaching load is kept constant. In such cases the evidence collected indicates that approximately the same average achievement can be expected from the pupils taught in large classes as from those taught in small classes. In other words, the results of this experiment indicate that there is no loss of efficiency caused by organizing a few large classes if the other work assigned to the teacher is such that the teaching load is not increased.

One should recognize that the results of this experiment should not be applied to the question of the size of class where increasing the size of class results in a distinct increase in the teaching load. The instruction which students receive is given partly in the class room and partly through written work and individual conferences. In such subjects as English composition, algebra, and science requiring laboratory work, it is customary with most teachers to require a large amount of written work. A teacher who gives instruction to five classes of 40 students each has a much heavier teaching load than the teacher who instructs five classes of 20 students each, unless he introduces compensating changes in the amount of written work, in the method of handling it, and in the number of individual conferences. In such cases the question of class size is so intimately connected with the method of instruction that we are not justified in drawing any inferences from an investigation in which the method of instruction was assumed to be the same for both types of classes.

CHAPTER V

SUGGESTIONS FOR EDUCATIONAL EXPERIMENTATION[1]

The two studies described in Chapters III and IV make very slight contributions to a scientific determination of the relation between the size of class and the efficiency of a school system. They do, however, seem to the writer of this chapter suggestive with reference to the procedure of educational experimentation. The causes of the failure of these studies to produce reliable and significant results have been mentioned in the two preceding chapters but they may be summarized under two general heads: *first,* failure to set up and maintain appropriate experimental conditions and, *second,* the lack of adequate instruments for measuring the achievements of pupils.

A prerequisite for setting up and maintaining appropriate experimental conditions is a complete analysis of the problem being studied. The various factors involved must be recognized by the experimenter and the possibility of any relations which may exist between these factors must also be considered. For example, in the high school many factors contribute to the achievements of students, or the educational output of the school. In addition to the size of class, which is the factor whose relationship to school efficiency was studied, it is necessary to recognize methods of instruction, the personality and enthusiasm of the teacher, the discipline of the class and of the school, the general spirit of the school, the general attitude of the community toward the school, the time of day when the class recites, the textbooks used, the equipment, including the building, the "spiritual composition" of the class, the general intelligence of the students, their nationality, their past experience, both in school and out of school, the stage of advancement reached in their education, and possibly even other factors. It appears likely that certain of these factors are interrelated. The size of class is likely to affect the enthusiasm of the teacher, particularly if it determines the in-

[1]This chapter has no counterpart in the report submitted by Mr. Stevenson. It is entirely the contribution of the present director of the Bureau of Educational Research.

structional load carried by the teacher. It also appears that some relationship exists between the size of class and methods of instruction, and between the size of class and discipline. The existence of a functional relationship between two or more factors makes it impossible under normal conditions to produce variations in one factor without, at the same time, causing changes in the others. Failure to analyze the problem sufficiently will frequently cause the results of an educational experimentation to have little significance, and consequently the time and money invested in the study will be largely wasted.

When one considers the total product of the educational process one cannot fail to become impressed with the inadequacy of our present educational tests as instruments for the measurement of the various elements of this product. In the study relating to the size of class in the elementary school a more elaborate group of tests might have been used, but even if this had been done, it does not appear likely that one would be justified in asserting that the total product of education had been measured. In the high school no standardized educational tests were used. It was attempted to have the achievements measured by means of a final examination and by the term "grade" given to the students. The writer of this chapter is not aware of the considerations which resulted in the decision not to use any of the standardized educational tests that have been devised in the field of high school subjects, but it is likely that this decision was due to the belief that none of the available educational tests were sufficiently satisfactory measuring instruments to justify their use in this investigation. The present writer is inclined to share this belief. Thus, we cannot escape the conclusion that at the present time we do not have available instruments for measuring the outcomes of teaching which permit reliable educational experimentation when it is necessary to measure the total product of instruction.

Incidentally, attention may be called to the fact that more consideration should be given to the errors involved in the data and to the effect of these errors upon the results of statistical calculations. For example, it is highly important to know what significance should be attached to a difference between two averages.

In view of the difficulties encountered in setting up and maintaining appropriate experimental conditions and in view of the im-

perfections and limitations of our present educational tests, it is not inappropriate to question the wisdom of undertaking such complex educational experimentation as has been described in this monograph. It is true that there are many educational problems which are highly important. For example, an increase in the size of class would result in a material reduction in the educational expenditures for instruction. A reliable scientific determination of the relationship existing between the size of class and school efficiency would be a valuable contribution, but it is doubtful whether such a determination of this relation is at the present time possible.

To the present writer, it appears highly important that those engaged in educational research should give very careful consideration to the sort of problems to which they devote their energies. It is, of course, necessary that pioneer work be done, and in studies of this type, it is not always possible to anticipate the limitations of one's procedure. As a result it may become necessary to "scrap" a project because of the defects that appear in the course of one's work. Such losses are unavoidable in extending the frontiers of educational research. When an investigation is not pioneer work an experimenter should determine the limitations of his procedure in advance, and unless it appears likely that information of considerable value will be secured in spite of the limitations the investigation should not be undertaken. The fact that a problem is important does not justify its study. Educational experimentation which involves the use of faulty method and faulty instruments not only fails to make adequate contributions to our educational progress, but, more important, it tends to reflect unfavorably upon the application of the methods of research to the field of education.

THE UNIVERSITY OF ILLINOIS

THE STATE UNIVERSITY

URBANA

DAVID KINLEY, PH.D., LL.D., *President*

The University Includes the Following Departments

THE GRADUATE SCHOOL

THE COLLEGE OF LIBERAL ARTS AND SCIENCES (Ancient and Modern Languages and Literatures; History, Economics, Political Science, Sociology, Philosophy, Psychology, Education; Mathematics; Astronomy; Geology; Physics; Chemistry; Botany, Zoology, Entomology; Physiology, Art and Design)

THE COLLEGE OF COMMERCE AND BUSINESS ADMINISTRATION (General Business, Banking, Insurance, Accountancy, Railway Administration, Foreign Commerce; Courses for Commercial Teachers and Commercial and Civic Secretaries)

THE COLLEGE OF ENGINEERING (Architecture; Architectural, Ceramic, Civil, Electrical, Mechanical, Mining, Municipal and Sanitary, Railway Engineering, and General Engineering Physics)

THE COLLEGE OF AGRICULTURE (Agronomy; Animal Husbandry; Dairy Husbandry; Horticulture and Landscape Gardening; Agricultural Extension; Teachers' Course; Home Economics)

THE COLLEGE OF LAW (Three-year and four-year curriculums based on two years and one year of college work respectively)

THE COLLEGE OF EDUCATION

THE CURRICULUM IN JOURNALISM

THE CURRICULUMS IN CHEMISTRY AND CHEMICAL ENGINEERING

THE SCHOOL OF RAILWAY ENGINEERING AND ADMINISTRATION

THE SCHOOL OF MUSIC (four-year curriculum)

THE LIBRARY SCHOOL (two-year curriculum for college graduates)

THE COLLEGE OF MEDICINE (in Chicago)

THE COLLEGE OF DENTISTRY (in Chicago)

THE SCHOOL OF PHARMACY (in Chicago; Ph.G. and Ph.C. curriculums)

THE SUMMER SESSION (eight weeks)

EXPERIMENT STATIONS AND SCIENTIFIC BUREAUS: U. S. Agricultural Experiment Station; Engineering and Experiment Station; State Laboratory of Natural History; State Entomologist's Office; Biological Experiment Station on Illinois River; State Water Survey; State Geological Survey; U. S. Bureau of Mines Experiment Station.

The library collections contain May 1, 1922, 523,230 volumes and 120,151 pamphlets. For catalogs and information address

THE REGISTRAR

Urbana, Illinois

BULLETINS OF THE BUREAU OF EDUCATIONAL RESEARCH, COLLEGE OF EDUCATION, UNIVERSITY OF ILLINOIS, URBANA, ILLINOIS.

		Price.
No. 1.	Buckingham, B. R. Bureau of Educational Research, Announcement, 1918-19	.15
No. 2.	First Annual Report	.25
No. 3.	Bamesberger, Velda C. Standard Requirements for Memorizing Literary Material	.50
No. 4.	Holley, Charles E. Mental Tests for School Use. (Out of print)	.50
No. 5.	Monroe, Walter S. Report of Division of Educational Tests for 1919-20	.25
No. 6.	Monroe, Walter S. The Illinois Examination	.50
No. 7.	Monroe, Walter S. Types of Learning Required of Pupils in the Seventh and Eighth Grades and in the High School	.15
No. 8.	Monroe, Walter S. A Critical Study of Certain Silent Reading Tests	.50
No. 9.	Monroe, Walter S. Written Examinations and Their Improvement. (In preparation)	.50
No. 10.	Bureau of Educational Research. Relation of Size of Class to School Efficiency	.50